



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/214945/>

Version: Published Version

Article:

Cull, Alyssa H., Kent, David G. and Warren, Alan J. (2024) Emerging genetic technologies informing personalized medicine in Shwachman-Diamond syndrome and other inherited BMF disorders. *Blood*. pp. 931-939. ISSN: 1528-0020

<https://doi.org/10.1182/blood.2023019986>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Emerging genetic technologies informing personalized medicine in SDS and other inherited bone marrow failure disorders

Tracking no: BLD-2023-019986-CR2

Alyssa Cull (University of York, United Kingdom) David Kent (University of York, United Kingdom)
Alan Warren (University of Cambridge, United Kingdom)

Abstract:

The ribosomopathy Shwachman-Diamond syndrome (SDS) is a rare autosomal recessive inherited bone marrow failure syndrome (IBMFS) caused by mutations in the Shwachman-Bodian-Diamond syndrome (SBDS) gene, that is associated with an increased risk of myeloid malignancy. Tracking how hematopoietic stem cell (HSC) clonal dynamics change over time, assessing whether somatic genetic rescue mechanisms affect these dynamics, and mapping out when leukemic driver mutations are acquired is important to understand which individuals with SDS may go on to develop leukemia. In this review, we will discuss how new technologies that allow researchers to map mutations at the level of single HSC clones are generating important insights into genetic rescue mechanisms and their relative risk for driving evolution to leukemia, and how these data can inform the future development of personalized medicine approaches in SDS and other IBMFSs.

Conflict of interest: COI declared - see note

COI notes: A.J.W. is a consultant for SDS Therapeutics. The remaining authors declare no competing interests.

Preprint server: No;

Author contributions and disclosures: A.H.C., D.G.K. and A.J.W. contributed equally to the text of this paper and A.H.C. created the figures.

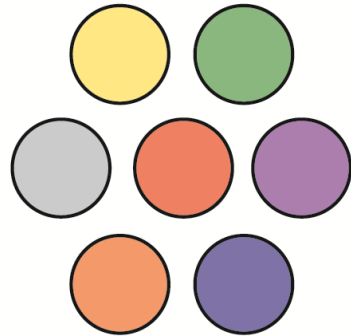
Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement:

Clinical trial registration information (if any):

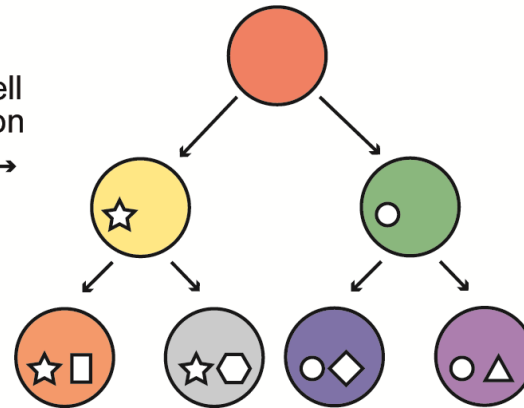
Emerging genetic technologies informing personalized medicine in Shwachman-Diamond Syndrome and other inherited bone marrow failure disorders

Sampled HSC Pool



Single Cell
Expansion
→
WGS

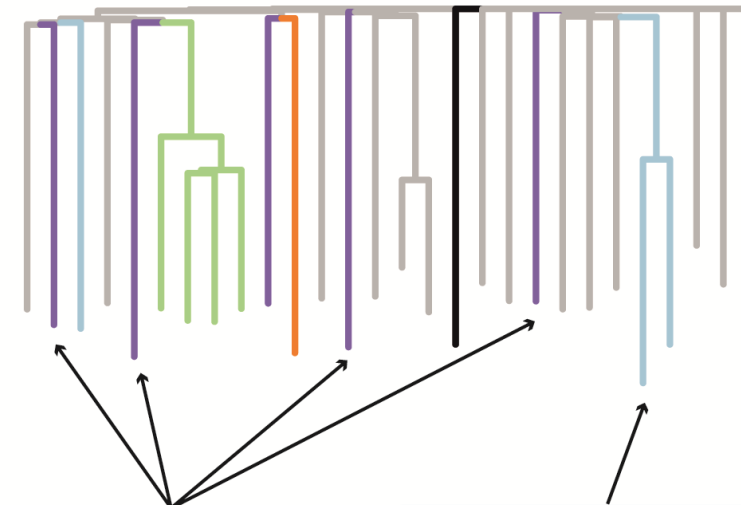
Phylogenetic Tree



Mapping HSC Relatedness in Disease

Building a better understanding of pathways affecting HSC selection

Shwachman-Diamond Syndrome



Maladaptive
TP53 Mutant Clones
Future MDS/AML?

Adaptive
EIF6 Mutations
Rescue Ribosome Defect

New technologies that allow researchers to map mutations at the level of single HSC clones are generating important insights into the pathways central to Shwachman-Diamond Syndrome disease biology.

Figure 1

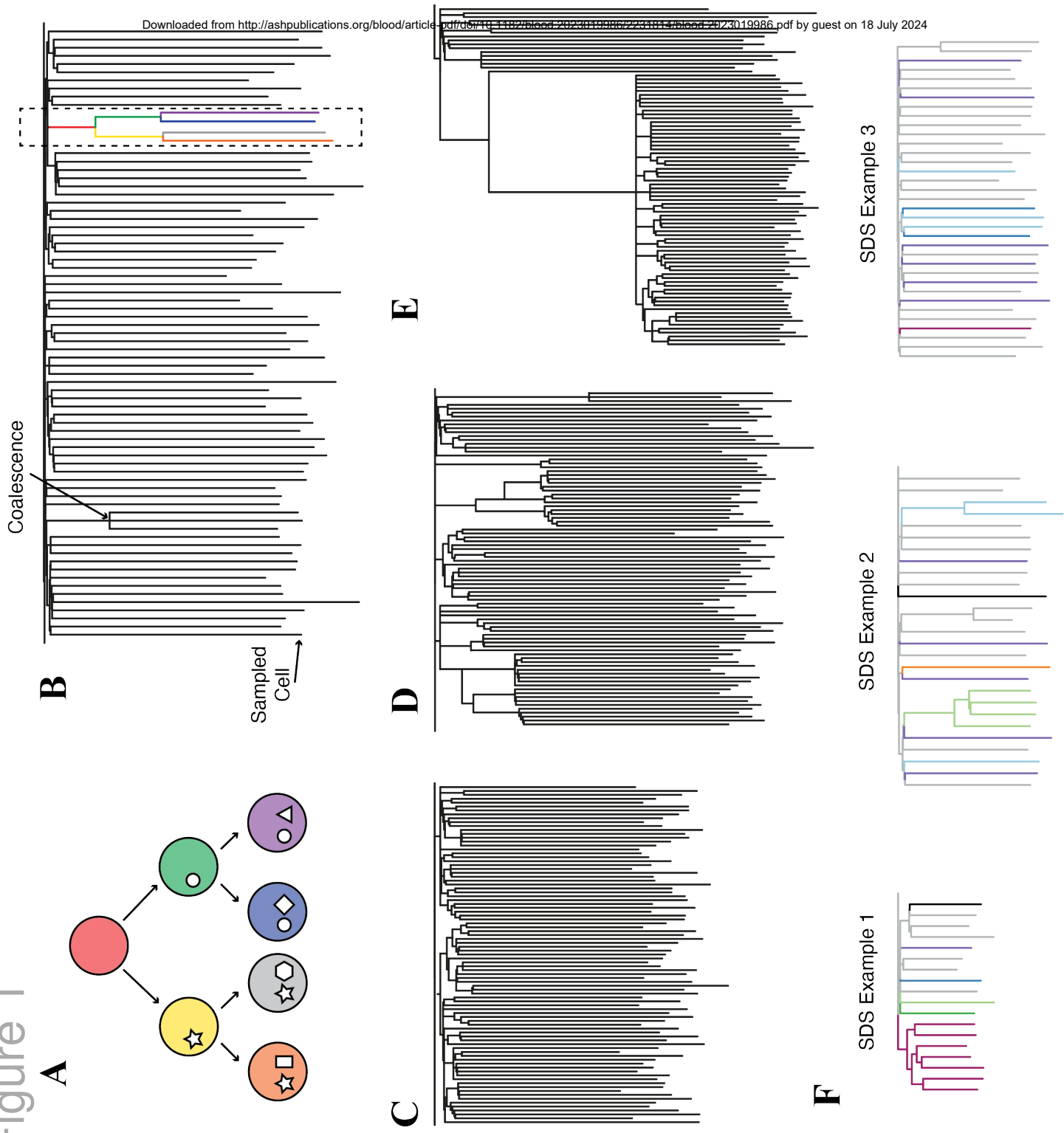
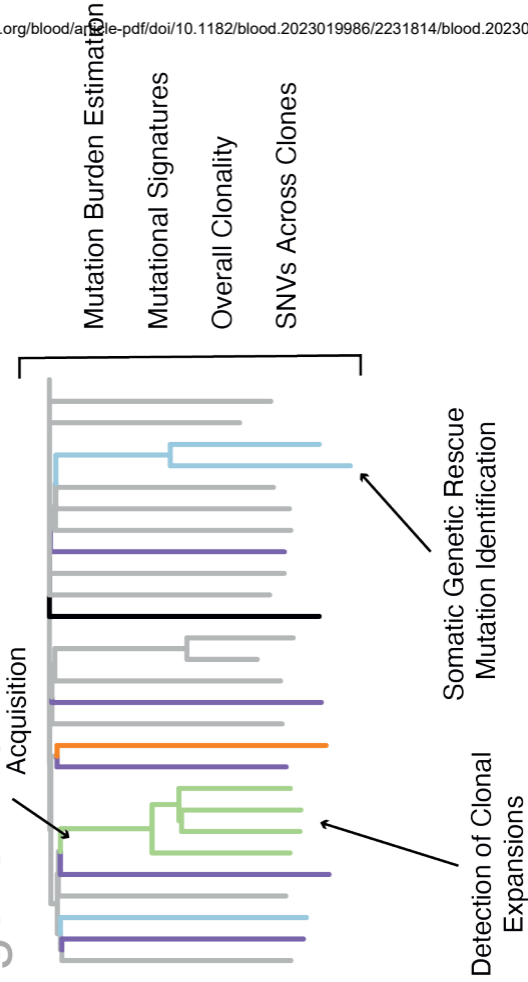


Figure 2
Timing of Mutation Acquisition



1 **Emerging genetic technologies informing personalized medicine in SDS and**
2 **other inherited bone marrow failure disorders**
3

4 Running Title: Applications of single-cell analyses to IBMFSs
5

6 Alyssa H. Cull¹, David G. Kent^{1*}, and Alan J. Warren^{2,3*}.
7

8 ¹York Biomedical Research Institute, Department of Biology, University of York, York, UK;

9 ²Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of
10 Cambridge, Cambridge Biomedical Campus, The Keith Peters Building, Hills Road, Cambridge CB2
11 0XY, UK.

12 ³Wellcome MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
13

14 * These authors contributed equally. Correspondence: david.kent@york.ac.uk and
15 ajw1000@cam.ac.uk
16

17 Text word count: 4590

18 Abstract word count: 135

19 Figure count: 2

20 Reference count: 79
21
22
23
24
25
26
27
28
29
30
31
32

33 **Abstract**

34 The ribosomopathy Shwachman-Diamond syndrome (SDS) is a rare autosomal recessive inherited
35 bone marrow failure syndrome (IBMFS) caused by mutations in the Shwachman-Bodian-Diamond
36 syndrome (*SBDS*) gene, that is associated with an increased risk of myeloid malignancy. Tracking
37 how hematopoietic stem cell (HSC) clonal dynamics change over time, assessing whether somatic
38 genetic rescue mechanisms affect these dynamics, and mapping out when leukemic driver mutations
39 are acquired is important to understand which individuals with SDS may go on to develop leukemia.
40 In this review, we will discuss how new technologies that allow researchers to map mutations at the
41 level of single HSC clones are generating important insights into genetic rescue mechanisms and their
42 relative risk for driving evolution to leukemia, and how these data can inform the future development
43 of personalized medicine approaches in SDS and other IBMFSs.

44

45 **Introduction**

46

47 Inherited bone marrow failure syndromes (IBMFSs) are heterogeneous diseases that typically present
48 with cytopenias and multi-organ dysfunction. The most common IBMFSs include Fanconi anemia
49 (FA), Diamond-Blackfan anemia (DBA), dyskeratosis congenita (DC), GATA2 deficiency,
50 SAMD9/9L-related syndromes, and Shwachman-Diamond syndrome (SDS). In addition to bone
51 marrow (BM) failure and ineffective hematopoiesis, the genetic drivers of these diseases often lead to
52 abnormalities in other tissues and patients are at a much-increased risk of developing a number of
53 cancers early in life, including leukemia.¹

54

55 The ribosomopathy SDS is a fascinating autosomal recessive BM failure disorder caused by defective
56 ribosome assembly.² In 90% of cases, SDS is caused by mutations in a single gene, Shwachman-
57 Bodian-Diamond syndrome (*SBDS*).^{3,4} The ribosome assembly defect in SDS results in a range of
58 highly variable clinical manifestations including neutropenia, poor growth, exocrine pancreatic
59 insufficiency, skeletal abnormalities, cognitive impairment and other organ dysfunction.² Individuals
60 with SDS are at higher risk of developing poor prognosis MDS/AML within the first few decades of
61 life.^{5-7,8} While most individuals are diagnosed as children, patients may still remain undiagnosed into
62 adulthood. Understanding why some SDS patients progress to MDS/AML whereas others do not is
63 key to developing preventative strategies and personalized therapeutic interventions.

64

65 Given that it is a monogenic disorder driving a fundamental biological defect in all cells, SDS
66 represents an excellent model disease to explore somatic genetic rescue mechanisms. For many years,
67 a key question in the field of SDS research remains: how does a germline ribosome defect associated
68 with a growth disadvantage lead to the development of leukemia? The risk of transformation for
69 individuals with SDS is high. In the French SDS registry, approximately 30% of patients developed a

70 myeloid malignancy by the age of 30.⁶ The cumulative 20-year incidence was 9.8% in the Italian SDS
71 registry,⁷ while in the North American SDS registry, 17% developed a myeloid cancer by 18 years of
72 age.^{8,9} Outcomes for patients who progress to myeloid neoplasms are very poor. In a recent
73 retrospective study, the three-year overall survival for patients who developed MDS or AML was
74 estimated to be 51% and 11%, respectively.⁸ Particularly in the case of SDS-related MDS, these
75 survival rates are worse than those observed in the non-SDS population.⁸ Understanding the genetic
76 changes that drive transformation to these aggressive blood cancers becomes of paramount
77 importance and while acquired biallelic *TP53* mutations are common in myeloid cancers in SDS, this
78 may not be the only route.^{8,10}

79

80 In this review, we detail how recently developed technologies have expanded our ability to map the
81 mutations involved in genetic rescue and leukemic transformation at the level of single hematopoietic
82 stem cell (HSC) clones. Understanding the range and type of genetic rescue mechanisms and their
83 relative risk for driving disease evolution of SDS to leukemia will help inform future drug design and
84 clinical decision making for IBMFSs more generally.

85

86 **Molecular basis of SDS**

87

88 Great strides have been made in understanding the molecular basis of SDS.^{2,11–13} Most individuals
89 with SDS carry germline mutations in the highly conserved *SBDS* gene encoding the SBDS protein
90 that licenses entry of the large ribosomal subunit into the actively translating pool of ribosomes.²
91 Approximately 200 proteins and 76 small nucleolar RNA molecules coordinate the assembly of 80
92 ribosomal proteins and four ribosomal RNAs into the small (40S) and large (60S) ribosomal subunits.
93 The initial steps in ribosome biogenesis occur in the nucleolus, where the precursors of the 40S and
94 60S subunits are produced via endonucleolytic cleavage of the 90S pre-ribosome.^{14–18} Upon export
95 into the cytoplasm, the nascent 40S and 60S ribosomal subunits complete the final steps in
96 maturation. One of the last cytoplasmic maturation events is eviction of the anti-association factor
97 eIF6 from the 60S ribosomal subunit by the concerted action of the GTPase elongation factor-like 1
98 (EFL1) and its cofactor SBDS.^{19–22} Removal of eIF6 from the intersubunit face of the nascent 60S
99 subunit is essential to allow 80S ribosome assembly as eIF6 physically blocks joining of the nascent
100 60S to the 40S ribosomal subunit. SBDS and EFL1 also act as general eIF6 release factors by
101 promoting the recycling of post-termination eIF6-bound 60S subunits back into active translation.²²

102

103 Located on chromosome 7q11, the *SBDS* gene is most commonly mutated within exon 2
104 (c.258>+2T>C and c.183_184TA>CT) due to a gene conversion event arising from recombination
105 between *SBDS* and an adjacent pseudogene *SBDSPI*, markedly reducing functional SBDS protein

106 expression.^{3,4} This reduction in SBDS expression impairs the tight coupling of EFL1 GTP hydrolysis
107 on the 60S subunit to eIF6 release, promoting eIF6 retention, impaired ribosome assembly and
108 reduced translation.²¹ In addition, mutations in genes such as *EFL1*,^{23–25} *DNAJC21*^{26,27} and *SRP54*^{28–30}
109 cause an SDS-like phenotype, consistent with their role in ribosome assembly (*EFL1*, *DNAJC21*) or
110 the recruitment of signal peptide-containing proteins to the endoplasmic reticulum for secretion or
111 membrane insertion (*SRP54*). All of these mutations lead to the striking defect in protein synthesis
112 observed in cells from individuals with SDS.^{2,22,25,31,32} Translational dysregulation due to loss of
113 SBDS function may contribute to the development of neutropenia in SDS.^{33–36}

114

115 **Somatic genetic rescue of defective ribosome assembly in SDS**

116

117 Somatic genetic rescue occurs in inherited disorders when an acquired somatic mutation or gross
118 chromosomal change partially or completely offsets the deleterious effects of a germline mutation.¹²
119 This can take the form of removal or correction of the mutation itself, or modification of other target
120 genes such that the pathways affected by the germline mutation are altered. While these changes
121 benefit the survival and function of individual cells, they may in fact have a neutral or detrimental
122 impact on further disease development. First reported in a patient with severe combined
123 immunodeficiency,³⁷ somatic genetic rescue mechanisms have been described in a number of IBMFSs
124 including FA, SDS, *SAMD9/9L* syndromes, DC and DBA.¹² In *SAMD9/9L* syndromes, gain-of-
125 function mutations in the *SAMD9* and *SAMD9L* genes lead to a general decrease in cellular
126 proliferative capacity.³⁸ In this context, one of the most common rescue mechanisms involves the loss
127 of chromosome 7 which reduces expression of the mutant allele in hematopoietic cells.^{38–43} While this
128 may be beneficial in the short-term, these changes have also been linked to an increased likelihood of
129 developing MDS and AML.^{44,45}

130

131 The p53 protein has been implicated as a key player in driving the clinical features of several
132 ribosomopathies including SDS, Treacher Collins syndrome, DC, DBA, North American Indian
133 Childhood Cirrhosis and 5q- syndrome^{36,46–51}. Impaired ribosome assembly stabilizes p53 via the
134 nucleolar surveillance pathway⁵². Thus, somatic mutations that reduce p53 activation could increase
135 cellular fitness either by directly inactivating *TP53* or by rescuing the underlying ribosome maturation
136 defect. In keeping with this hypothesis, significantly more *TP53* loss-of-function mutations were
137 detected in patients with biallelic *SBDS* mutations compared to those with one or no *SBDS* mutations
138 or other forms of neutropenia.⁵³ Mutations in eIF6 suppress the fitness defect of yeast cells lacking the
139 SBDS homologue.¹⁹ It was therefore striking to find that in SDS, hematopoietic cells' common
140 compensatory mechanisms include diverse mosaic somatic genetic events such as missense mutations
141 in *EIF6*, reciprocal chromosomal translocation or interstitial deletions of chromosome 20 (del(20)q)

142 that encompass the *EIF6* gene^{11,54}, either reducing eIF6 expression or disrupting its interaction with
143 the 60S subunit to confer a selective advantage over non-modified cells^{31,32}. Somatic genetic rescue
144 mutations have also been described in *EFL1* mutated patients.⁵⁵ If acquired in HSCs, the contribution
145 of an individual stem cell to hematopoiesis alters over time if these somatic genetic rescue mutations
146 and chromosomal changes enhance clonal fitness. Understanding the dynamic behavior of clones
147 carrying specific somatic mutations is therefore vital to determine which patients are at greater risk of
148 developing leukemia and how these blood cancers may differ from leukemia that occurs later in
149 adulthood.

150 **New approaches to addressing unanswered questions**

151 Despite a general understanding of which somatic rescue mutations are preferentially acquired in
152 *SBDS*-mutated hematopoietic cells, little is known about the nature and dynamics of those mutations
153 and whether multiple mutations occur in the same or different clones.^{10,32,56} Recent efforts using new
154 technologies have been undertaken to tie mutations to specific clones and to assess clonal dynamics.
155 While next generation sequencing (NGS) technologies and targeted sequencing panels have
156 revolutionized diagnostic medicine within the field of hematology, the high error rate of traditional
157 NGS limits its use in the identification of mutant subclones, especially at lower allele burdens.⁵⁷
158 While digital droplet PCR is a sensitive and cost-effective way of monitoring low variant allele
159 fraction (VAF) clones, it requires prior knowledge of mutations.⁵⁸ In the last decade, duplex
160 sequencing approaches have enabled the identification of very low frequency somatic mutations from
161 heterogeneous cell populations by using specialized tags and sequencing both strands of DNA to
162 greatly reduce the error rate.^{57,59} This technology is able to reliably detect a single somatic mutation in
163 $>10^7$ sequenced bases.⁶⁰ An even more recent approach (NanoSeq) further reduced error rates down to
164 <5 errors per billion bases sequenced.⁶¹ Once generated, whole genome sequencing (WGS) data can
165 then be used to interrogate features such as intratumoral heterogeneity.⁶² In addition to detection of
166 single nucleotide variants (SNVs), technologies such as optical genome mapping provide detailed
167 information on structural variants (SVs).⁶³ While these advances have allowed for single-molecule
168 detection of SNVs, high resolution mapping of SVs, and potential longitudinal monitoring of very low
169 VAF mutant clones, none of these bulk sequencing technologies can be used to identify new clones
170 with extremely low VAF mutations or to comprehensively map the origins of, or relationships
171 between, different HSC clones.

172
173 One of the few ways in which unperturbed human HSC relatedness can be studied involves the use of
174 naturally occurring somatic mutations as a genetic barcode.^{64,65} In brief, this technique tracks the
175 spontaneously acquired SNVs that occur in an organism through normal cellular processes that lead to
176 DNA damage (e.g., cell division, mutagen exposure, etc). Human HSCs acquire these mutations

177 linearly over time, at a rate of approximately 14-18 SNVs and 0.65-0.77 insertions/deletions (indels)
178 each year.⁶⁵⁻⁶⁷ The majority of these changes are neutral with no effect on clonal fitness or HSC
179 function and the patterns of shared and unique SNVs can be used to create phylogenetic trees (Figure
180 1A-B). These diagrams visually depict HSC clonal relationships, with branch points (termed
181 ‘coalescences’) representing historic stem cell self-renewal divisions where one HSC gave rise to two
182 daughter HSCs, each of whose progeny can be distinguished through mutational profiling of shared
183 and unique mutations (Figure 1B). This approach has recently been used to study both normal and
184 diseased states.^{56,64,67-71} In healthy adults under the age of 70, the HSC pool is diverse and highly
185 polyclonal, with many HSCs contributing to hematopoiesis (Figure 1C).⁶⁷ In elderly individuals over
186 the age of 70, clonal expansions become much more prevalent with a relatively limited pool of stem
187 cells supporting blood production (Figure 1D).⁶⁷ Tree structures become even more skewed towards
188 dominant clones in the context of blood cancers (Figure 1E).⁶⁸ In addition to detailing HSC
189 relatedness, clonal information can also be used to estimate overall mutation burden and stem cell
190 numbers, identify specific mutational processes that might drive SNV acquisition (mutational
191 signatures), and detect and estimate the timing of specific and sometimes unexpected driver mutations
192 in expanded clones (Figure 2).⁷¹ Since the above described phylogenetic approach relies on creating
193 numerous WGS libraries per patient to map and track large numbers of competing clones, it can
194 become prohibitively expensive. Additionally, limitations to this approach exist. For example, the
195 assay relies on the expansion of single hematopoietic stem and progenitor cells (HSPCs), either in
196 liquid culture or in Methocult medium, to provide enough genetic material for sequencing. The
197 efficiency of colony growth in liquid culture from FACS-sorted HSPCs may be ~70%.⁶⁷ However,
198 depending on the cell source (sorted versus enriched), this number can be lower since more
199 differentiated cells will not produce colonies. This may also be disease- or patient-specific, with some
200 IBMFS samples growing poorly *in vitro* and it is possible that some rare mutant clones will be missed
201 as a result. As such, alternative ways to map clonal dynamics have been pursued, including the use of
202 mitochondrial DNA mutations rather than genome-wide SNVs as molecular barcodes.⁷² Another
203 option is to use single-cell long-read WGS,⁷³ but a range of technical challenges exist that currently
204 limit its use in establishing lineage relationships.⁷⁴ Therefore, while targeted sequencing panels,
205 especially those utilizing high sensitivity duplex sequencing, will become a powerful clinical tool for
206 tracking mutations (and their respective clones) in patients, if clonal origins and/or lineage
207 relationships are required, then whole genome approaches remain the only robust method.

208

209 **Mapping relationships between HSCs in SDS**

210

211 Using the powerful method of SNV profiling and HSC phylogeny building in the context of SDS,
212 Machado *et al.* recently mapped HSC relatedness and assessed somatic genetic rescue mechanisms in

213 samples taken from ten individuals with biallelic germline loss-of-function mutations in *SBDS*.⁵⁶
214 WGS was performed on DNA isolated from 323 individually expanded single hematopoietic stem and
215 progenitor cells (HSPCs) obtained from either the peripheral blood (PB) or bone marrow (BM) of
216 patients. For seven of the ten patients (age range 4-33 years old), substantial clonal expansions were
217 observed, as evidenced by clustering of the sampled HSPCs into large clades (Figure 1F).
218 Phylogenetic tree structures like this are typically only observed much later in life⁶⁷ which suggests
219 that for patients with SDS, HSC diversity and clonal architecture is profoundly altered within the first
220 few decades of life. Whether reduced polyclonality is observed in the HSC pools of patients with
221 other IBMFSs remains to be seen, especially since this feature was not uniformly observed in all
222 patients studied by Machado *et al.*

223

224 The mutational information generated by WGS also identifies specific somatic mutations acquired by
225 individual ancestral HSCs. Machado *et al.* found that *TP53* was the most commonly mutated gene in
226 individual HSCs which accorded with data from larger cohorts of SDS patients reported by Kennedy
227 *et al.*³² and Xie *et al.*,⁵³ where 45% and 48% of SDS patients harbored a *TP53* mutation, respectively.
228 The Kennedy study also reported that ~60% of patients carried *EIF6* mutations, a finding reinforced
229 by both the Machado *et al.* and Tan *et al.* studies.³¹ Beyond these most common mutations, Machado
230 *et al.* also detected SNVs in genes reported as recurrently mutated in SDS patients (e.g. *PRPF8*),
231 while also for the first time in SDS identifying mutations in ribosomal protein genes *RPL5* (encoding
232 uL18) and *RPL22* (encoding eL22). Although still to be established, these mutations may decrease
233 p53 activation by reducing signaling through the nucleolar stress pathway, potentially diminishing the
234 drive to select *TP53* mutations, thereby lowering the risk of disease progression. Together, this
235 approach allows researchers to comprehensively map the landscape of potential mutations that rescue
236 the *SBDS*-associated ribosome deficiency, but do not compromise cell viability or lead to frequent
237 leukemic transformation. Since every single cell in SDS is under evolutionary pressure to rescue the
238 *SBDS*-dependent ribosome deficiency, the exploration of possible genetic rescue mechanisms across
239 multiple patients is extensive. Mutations that are under positive selection in SDS lead to improved
240 cellular fitness and the persistence of clones bearing those mutations.

241

242 Whether or not individual clones need single or multiple somatic genetic rescue events and whether
243 mutations involved in clonal hematopoiesis (CH) co-occur with rescue mutations is less clear. Across
244 both the Kennedy and Machado studies, CH itself was detected in up to 72% of SDS patients.^{32,56}
245 Most of these expansions were associated with somatic genetic rescue mutations rather than typical
246 age-related CH mutations in genes such as *DNMT3A*, *TET2* and *ASXL1*.^{32,56} By sequencing serial
247 blood samples taken from 49 SDS patients with detectable CH, Kennedy *et al.* demonstrated that
248 many of these driver mutations were consistently seen across multiple samples and the VAFs of these
249 mutations remained low and stable over time.³² While this is a powerful approach to follow clonal

250 evolution, it is limited by the availability of longitudinal samples, especially those taken prior to SDS
251 diagnosis. By using somatic mutations as natural barcodes, the tree building approach skirts this
252 pitfall because it concomitantly reports the historical origins of the mutation (and therefore the growth
253 rates over the lifetime of the individual) as the sequencing information collected from HSPCs reflects
254 the mutational state of ancestral HSCs. Regarding co-occurrence of CH mutations and gene rescue
255 events, the Machado *et al.* study provides evidence in single clones that points to mutual exclusivity
256 in most cases (including SDS-related or CH-related mutations such as those affecting *DNMT3A*,
257 *ASXL1*, *TET2* and *RUNX1*). The two exceptions to this were (1) a clone with a combination of *TET2*
258 and *TP53* mutations and (2) a patient with biallelic *TP53* mutations who went on to develop MDS
259 with acquisition of multiple copy number alterations and mutations. This, in combination with the
260 observation that the individual genome-wide mutation burden was not increased in *TP53* mono-allelic
261 mutant SDS patients, suggests that heterozygous somatic genetic rescue events were sufficient to
262 restore cellular fitness and compensate for the SBDS deficiency. This conclusion is also consistent
263 with the study by Kennedy *et al.* where they first used bulk whole exome sequencing to identify
264 patients with multiple *EIF6* and *TP53* mutations and then employed a single cell DNA genotyping
265 approach to show that the vast majority of these mutations were acquired alone in distinct clones.³² In
266 both studies, the one important exception was biallelic *TP53* mutations that were associated with
267 genomic instability and progression to myeloid malignancy.^{32,56}

268

269 **Timing of mutation acquisition**

270

271 One of the most unique aspects of the phylogenetic tree building approach is its ability to estimate
272 when in an individual's lifetime specific driver mutations were acquired and when specific clonal
273 expansions began. Previous approaches to timing SNVs have involved serial sampling and sequencing
274 to look for earliest detection of low variant allele fraction (VAF) mutations. There are several
275 limitations to this approach. If standard bulk sequencing is used, mutations towards the lower end of
276 detection (e.g. mutations at low VAF) are likely to be missed. Therefore, a single cell approach using
277 serial samples as performed by Kennedy *et al.* would be the best option to yield high quality
278 information.³² However, this approach is constrained by the availability of serial samples. When using
279 the phylogeny approach, a single sample time point captures information about sequential SNV
280 acquisition in ancestral HSCs over time. Williams *et al.* were able to demonstrate that *JAK2*^{V617F}
281 mutations in myeloproliferative neoplasms were acquired early in life (between 33 weeks of gestation
282 and 11 years old) despite the absence of disease symptoms for many decades⁶⁸ and Machado *et al.*
283 were similarly able to estimate that clonal expansions with acquired driver mutations began as early as
284 *in utero* up to 12 years of age in SDS patients. Moreover, in individuals with SDS where there is
285 strong selective pressure to rescue the ribosomal defect, most mutations could be traced back to the

286 first decade of life and, interestingly, there were no observed patterns in terms of when mutations in
287 specific genes were acquired (including several CH-related mutations). While the sampling method
288 itself is biased toward clones that have had a longer time to expand, it is worth noting that studies in
289 normal aging have shown a much more protracted period of driver mutation acquisition compared to
290 SDS patients.⁶⁷

291

292 **Implications for patient monitoring**

293

294 Given the high risk of progression to myeloid cancers for patients with SDS and the dismal outcomes
295 of patients with pre-existing *TP53* mutations even after allogeneic bone marrow transplant⁷⁵, the
296 retrospective study of patients who have developed MDS/AML is key to developing new monitoring
297 and treatment strategies. In both the Kennedy *et al.* (whole exome sequencing) and Machado *et al.*
298 (WGS) studies, several driver mutations were identified in SDS patients that can now be investigated
299 for their impact on disease progression. Together, these studies identified recurrent mutations in
300 *TP53*, *EIF6*, *CSNK1A1*, *PRPF8*, *RPL5* and *RPL22* and as well as recurrent chromosomal
301 abnormalities in chromosomes 3, 5, 7 and 20.^{32,56} While mutations in genes such as *EIF6*, *CSNK1A1*,
302 and *PRPF8* were not found to be associated with blood cancer development, homozygous *TP53*
303 mutations were observed in all those SDS patients included in the study who had progressed to MDS
304 or AML compared to those who had not. Notably, patients in the Kennedy study who had not yet
305 progressed to cancer still had detectable heterozygous *TP53* mutations and this accords with no
306 increased mutation burden in *TP53* heterozygous clones reported by Machado *et al.* Together, these
307 data indicate that presence of a heterozygous *TP53* mutation alone does not predict that an individual
308 will go on to develop MDS/AML and suggests that this parameter by itself is of limited clinical value
309 in determining appropriateness for transplantation.

310

311 Also of interest, one of the patients in the Machado *et al.* study who had acquired biallelic *TP53*
312 mutations in the context of MDS transformation, showed substantial genome instability as evidenced
313 by a multiple copy number aberrations including ch5⁻, 6p⁻, 11⁺, 18⁻ and X⁻. This was further associated
314 with an increased overall mutation burden, more than double that expected for age. *TP53*-mutated
315 clones contained a higher total mutation burden than expected based on age and an increased
316 proportion of the mutational signature ‘SBS1’, linked to rapid cell division, which notably was
317 substantially lower in clones that pre-dated the most recently branched HSC clones, suggesting that
318 the increase in SBS1 signature-related mutations was associated with the onset of MDS. In addition,
319 the estimated clonal growth rate of 5200% per year for the biallelic *TP53* mutant clones (doubling
320 roughly every 2 months), reinforces the requirement for frequent surveillance of SDS patients.⁵⁶ NGS
321 results from BM and peripheral blood appear to be highly concordant, at least in the context of

322 hematological malignancy.⁷⁶ These data support the use of a combination of array comparative
323 genomic hybridization and myeloid NGS performed on serial peripheral blood samples (perhaps 6-12
324 monthly), in conjunction with frequent (3-4 monthly) full blood count sampling as a potential
325 surveillance strategy for SDS patients. That being said, the detection of mutations in IBMFS patients
326 with severe peripheral blood cytopenias may prove to be challenging and requires further
327 investigation. Bone marrow examination would be advisable if there is a significant alteration in
328 blood counts or if rising VAFs for specific variants (especially in the *TP53* gene) were detected.

329

330 **Future perspectives**

331

332 The advent of the phylogeny building approach has begun to open new areas of research into the
333 clonal dynamics of mutated HSCs during CH and disease progression, revealing the most common
334 recurrent mutations and their relative risk with respect to clonal dominance and/or disease evolution.
335 While the above section details its powerful use in a specific inherited bone marrow failure disorder,
336 SDS, to comprehensively map the entire complement of genetic rescue mechanisms and how these
337 might impact clinical evolution, the technology can be readily applied to other IBMFS and also to
338 diseases driven entirely by somatic mutations such as VEXAS (vacuoles, E1 enzyme, X-linked,
339 autoinflammatory, somatic) syndrome⁷⁷. Once mapped, these data can be used in two main ways for
340 clinical benefit: (1) to aid the design of targeted panels that cheaply and efficiently capture the
341 mutational profile of an individual patient; and (2) to identify “adaptive” and “maladaptive” rescue
342 mechanisms that may influence clinical outcomes. In both cases, this information can inform clinical
343 decision-making and risk stratification. This is critically important since retrospective analysis of data
344 in the North American SDS registry suggest a median survival of 7.7 years after MDS diagnosis and 1
345 year after AML diagnosis.⁸

346

347 The idea that inherited disorders that induce strong selective pressure will end up as composites of
348 hundreds to thousands of differently mutated cells prompts us to also imagine how to use this
349 information to help in clinical management. Could clones carrying adaptive mutations that restore
350 ribosome homeostasis and are associated with low risk of transformation be “boosted” in some
351 manner or those with high risk of transformation be targeted? Given the high frequency of *TP53*
352 mutation acquisition in SDS HSCs, this prompts the challenge of how best to detect *TP53* mutations
353 (particularly biallelic *TP53* mutations) in the clinical setting to inform decision-making (e.g.,
354 proceeding to transplant). Similarly, when high depth, high sensitivity methods permit reliable
355 detection of biallelic *TP53* loss in clinical assays, how might this impact prognosis? One clear method
356 would be to sample patients longitudinally to understand whether individually mutated clones are
357 growing rapidly over time, but this approach could be costly and time-consuming for the researcher

358 and the patient and may be further confounded by new clones emerging with a stronger selective
359 growth advantage. A potential route to avoid this latter challenge is to develop dedicated
360 SDS/myeloid panels that include all the mutations described above in targeted panel design so that
361 newly emerging clones bearing key SDS driver mutations can be tracked.

362

363 One thing is clear from the Kennedy and Machado studies - the complexity of the mutational
364 landscape in SDS is enormous. This presents particular challenges for therapeutic approaches such as
365 corrective gene therapy for a number of reasons. Firstly, the somatic rescue mutations that have arisen
366 to compensate for the absence of SBDS are already present from a very early age. Initiatives like the
367 Newborn Genomes Programme (<https://www.genomicsengland.co.uk/initiatives/newborns>) are
368 investigating the potential benefits of early patient diagnosis. In the UK, the Generation Study will
369 sequence the genomes of 100,000 newborn babies in England in partnership with the National Health
370 Service to understand whether improvements can be made in diagnosing and treating genetic
371 conditions. The full complement of causative genes in SDS patients are included in this study. In
372 future, it also might become possible to screen for cancer-associated mutations as these may be
373 acquired in utero.^{56,68} Secondly, correcting a specific genetic change (e.g., mutant *TP53*) will still
374 leave behind a wide range of other mutated HSC clones, so any gene therapy “fix” may only be
375 temporary. One approach might involve selectively steering clonal dynamics to promote the
376 outgrowth of adaptive clones (e.g., specific *EIF6*, *RPL22* or *RPL5* mutations, etc.) or restrict the
377 development, growth or persistence of biallelic (or indeed heterozygous) *TP53* mutant clones.

378

379 The recent new biological insights into the molecular mechanisms underpinning SDS and evolution to
380 MDS and AML generated using these new technologies highlight the importance of routinely
381 integrating serial next generation sequencing and microarray analysis into hematological surveillance
382 strategies for patients. Tools to study the relative competitive ability and potential leukemogenicity of
383 individual mutant clones are not currently well-developed and the WGS approach is not practical to
384 undertake at a large scale. Targeted sequencing panels may be useful to detect and track the most
385 common genes/mutations and variants when they reach a clonal burden of sufficient size, but
386 understanding the mechanism of each mutation will require more sophisticated genetic tools and
387 functional studies. As most myeloid malignancies in SDS involve the acquisition of biallelic *TP53*
388 mutations, there is an urgent need to develop robust tools to detect such clones at an early stage, at
389 very low allele frequency, prior to progression to overt MDS or AML, to allow for early consideration
390 of HSC transplantation. In this context, it also becomes vital to understand whether early transplant
391 for SDS patients who have acquired biallelic *TP53* mutations will improve the otherwise poor long
392 term outcomes.^{8,75,78,79} A better understanding of the consequences of somatic mutation acquisition on
393 clonal dynamics in SDS may also guide the development of therapeutics that mimic the effects of
394 adaptive mutations in restoring ribosome homeostasis, while avoiding the potential leukemic risk of

395 heterozygous *TP53* mutation acquisition and the leukemogenic potential of such clones if they arise.
 396 More generally, the identification of somatic rescue mutations across IBMFSs offers the potential to
 397 identify new targets for the design of disease modifying therapeutics.

398

399 References

- 400 1. Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in the National Cancer Institute inherited
 401 bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica*.
 402 2018;103(1):30–39.
- 403 2. Warren AJ. Molecular basis of the human ribosomopathy Shwachman-Diamond syndrome. *Adv.*
 404 *Biol. Regul.* 2018;67:109–127.
- 405 3. Boocock GRB, Morrison JA, Popovic M, et al. Mutations in SBDS are associated with
 406 Shwachman-Diamond syndrome. *Nat. Genet.* 2003;33(1):97–101.
- 407 4. Woloszynek JR, Rothbaum RJ, Rawls AS, et al. Mutations of the SBDS gene are present in most
 408 patients with Shwachman-Diamond syndrome. *Blood*. 2004;104(12):3588–3590.
- 409 5. Dror Y, Durie P, Ginzberg H, et al. Clonal evolution in marrows of patients with Shwachman-
 410 Diamond syndrome: a prospective 5-year follow-up study. *Exp. Hematol.* 2002;30(7):659–669.
- 411 6. Donadieu J, Fenneteau O, Beaupain B, et al. Classification of and risk factors for hematologic
 412 complications in a French national cohort of 102 patients with Shwachman-Diamond syndrome.
 413 *Haematologica*. 2012;97(9):1312–1319.
- 414 7. Cesaro S, Pegoraro A, Sainati L, et al. A Prospective Study of Hematologic Complications and
 415 Long-Term Survival of Italian Patients Affected by Shwachman-Diamond Syndrome. *J. Pediatr.*
 416 2020;219:196–201.e1.
- 417 8. Myers KC, Furutani E, Weller E, et al. Clinical features and outcomes of patients with
 418 Shwachman-Diamond syndrome and myelodysplastic syndrome or acute myeloid leukaemia: a
 419 multicentre, retrospective, cohort study. *Lancet Haematol.* 2020;7(3):e238–e246.
- 420 9. Furutani E, Liu S, Galvin A, et al. Hematologic complications with age in Shwachman-Diamond
 421 syndrome. *Blood Adv.* 2022;6(1):297–306.
- 422 10. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome
 423 stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat. Med.*
 424 2020;26(10):1549–1556.
- 425 11. Valli R, Pressato B, Marletta C, et al. Different loss of material in recurrent chromosome 20
 426 interstitial deletions in Shwachman-Diamond syndrome and in myeloid neoplasms. *Mol.*
 427 *Cytogenet.* 2013;6(1):56.
- 428 12. Revy P, Kannengiesser C, Fischer A. Somatic genetic rescue in Mendelian haematopoietic
 429 diseases. *Nat. Rev. Genet.* 2019;20(10):582–598.
- 430 13. Reilly CR, Shimamura A. Predisposition to myeloid malignancies in Shwachman-Diamond
 431 syndrome: biological insights and clinical advances. *Blood*. 2023;141(13):1513–1523.
- 432 14. Dragon F, Gallagher JEG, Compagnone-Post PA, et al. A large nucleolar U3 ribonucleoprotein
 433 required for 18S ribosomal RNA biogenesis. *Nature*. 2002;417(6892):967–970.
- 434 15. Grandi P, Rybin V, Bassler J, et al. 90S pre-ribosomes include the 35S pre-rRNA, the U3
 435 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol.*
 436 *Cell*. 2002;10(1):105–115.
- 437 16. Kornprobst M, Turk M, Kellner N, et al. Architecture of the 90S Pre-ribosome: A Structural
 438 View on the Birth of the Eukaryotic Ribosome. *Cell*. 2016;166(2):380–393.
- 439 17. Chaker-Margot M, Barandun J, Hunziker M, Klinge S. Architecture of the yeast small subunit
 440 processome. *Science*. 2017;355:6321.
- 441 18. Sun Q, Zhu X, Qi J, et al. Molecular architecture of the 90S small subunit pre-ribosome. *Elife*.
 442 2017;6:e22086.
- 443 19. Menne TF, Goyenechea B, Sánchez-Puig N, et al. The Shwachman-Bodian-Diamond syndrome
 444 protein mediates translational activation of ribosomes in yeast. *Nat. Genet.* 2007;39(4):486–495.
- 445 20. Finch AJ, Hilcenko C, Basse N, et al. Uncoupling of GTP hydrolysis from eIF6 release on the

- 446 ribosome causes Shwachman-Diamond syndrome. *Genes Dev.* 2011;25(9):917–929.
- 447 21. Weis F, Giudice E, Churcher M, et al. Mechanism of eIF6 release from the nascent 60S
448 ribosomal subunit. *Nat. Struct. Mol. Biol.* 2015;22(11):914–919.
- 449 22. Jaako P, Faille A, Tan S, et al. eIF6 rebinding dynamically couples ribosome maturation and
450 translation. *Nat. Commun.* 2022;13(1):1562.
- 451 23. Stepensky P, Chacón-Flores M, Kim KH, et al. Mutations in EFL1, an SBDS partner, are
452 associated with infantile pancytopenia, exocrine pancreatic insufficiency and skeletal anomalies
453 in a Shwachman-Diamond like syndrome. *J. Med. Genet.* 2017;54(8):558–566.
- 454 24. Tan QK-G, Cope H, Spillmann RC, et al. Further evidence for the involvement of EFL1 in a
455 Shwachman-Diamond-like syndrome and expansion of the phenotypic features. *Cold Spring
456 Harb Mol Case Stud.* 2018;4(5):a003046.
- 457 25. Tan S, Kermasson L, Hoslin A, et al. EFL1 mutations impair eIF6 release to cause Shwachman-
458 Diamond syndrome. *Blood.* 2019;134(3):277–290.
- 459 26. Tummala H, Walne AJ, Williams M, et al. DNAJC21 Mutations Link a Cancer-Prone Bone
460 Marrow Failure Syndrome to Corruption in 60S Ribosome Subunit Maturation. *Am. J. Hum.
461 Genet.* 2016;99(1):115–124.
- 462 27. Dhanraj S, Matveev A, Li H, et al. Biallelic mutations in DNAJC21 cause Shwachman-Diamond
463 syndrome. *Blood.* 2017;129(11):1557–1562.
- 464 28. Bellanné-Chantelot C, Schmaltz-Panneau B, Marty C, et al. Mutations in the SRP54 gene cause
465 severe congenital neutropenia as well as Shwachman-Diamond-like syndrome. *Blood.*
466 2018;132(12):1318–1331.
- 467 29. Carapito R, Konantz M, Paillard C, et al. Mutations in signal recognition particle SRP54 cause
468 syndromic neutropenia with Shwachman-Diamond-like features. *J. Clin. Invest.*
469 2017;127(11):4090–4103.
- 470 30. Juaire KD, Lapouge K, Becker MMM, et al. Structural and Functional Impact of SRP54
471 Mutations Causing Severe Congenital Neutropenia. *Structure.* 2021;29(1):15–28.e7.
- 472 31. Tan S, Kermasson L, Hilcenko C, et al. Somatic genetic rescue of a germline ribosome assembly
473 defect. *Nat. Commun.* 2021;12(1):5044.
- 474 32. Kennedy AL, Myers KC, Bowman J, et al. Distinct genetic pathways define pre-malignant
475 versus compensatory clonal hematopoiesis in Shwachman-Diamond syndrome. *Nat. Commun.*
476 2021;12(1):1334.
- 477 33. In K, Zaini MA, Müller C, et al. Shwachman-Bodian-Diamond syndrome (SBDS) protein
478 deficiency impairs translation re-initiation from C/EBP α and C/EBP β mRNAs. *Nucleic Acids
479 Res.* 2016;44(9):4134–4146.
- 480 34. Rawls AS, Gregory AD, Woloszynek JR, Liu F, Link DC. Lentiviral-mediated RNAi inhibition
481 of Sbds in murine hematopoietic progenitors impairs their hematopoietic potential. *Blood.*
482 2007;110(7):2414–2422.
- 483 35. Orelia C, Verkuijlen P, Geissler J, van den Berg TK, Kuijpers TW. SBDS expression and
484 localization at the mitotic spindle in human myeloid progenitors. *PLoS One.* 2009;4(9):e7084.
- 485 36. Zambetti NA, Bindels EMJ, Van Strien PMH, et al. Deficiency of the ribosome biogenesis gene
486 Sbds in hematopoietic stem and progenitor cells causes neutropenia in mice by attenuating
487 lineage progression in myelocytes. *Haematologica.* 2015;100(10):1285–1293.
- 488 37. Hirschhorn R, Yang DR, Israni A, Huie ML, Ownby DR. Somatic mosaicism for a newly
489 identified splice-site mutation in a patient with adenosine deaminase-deficient immunodeficiency
490 and spontaneous clinical recovery. *Am. J. Hum. Genet.* 1994;55(1):59–68.
- 491 38. Buonocore F, Kühnen P, Suntharalingham JP, et al. Somatic mutations and progressive
492 monosomy modify SAMD9-related phenotypes in humans. *J. Clin. Invest.* 2017;127(5):1700–
493 1713.
- 494 39. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder,
495 MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat. Genet.*
496 2016;48(7):792–797.
- 497 40. Tesi B, Davidsson J, Voss M, et al. Gain-of-function SAMD9L mutations cause a syndrome of
498 cytopenia, immunodeficiency, MDS, and neurological symptoms. *Blood.* 2017;129(16):2266–
499 2279.
- 500 41. Pastor VB, Sahoo SS, Boklan J, et al. Constitutional SAMD9L mutations cause familial

- 501 myelodysplastic syndrome and transient monosomy 7. *Haematologica*. 2018;103(3):427–437.
- 502 42. Wong JC, Bryant V, Lamprecht T, et al. Germline SAMD9 and SAMD9L mutations are
503 associated with extensive genetic evolution and diverse hematologic outcomes. *JCI Insight*.
504 2018;3(14):e121086.
- 505 43. Sahoo SS, Pastor VB, Goodings C, et al. Clinical evolution, genetic landscape and trajectories of
506 clonal hematopoiesis in SAMD9/SAMD9L syndromes. *Nat. Med.* 2021;27(10):1806–1817.
- 507 44. Davidsson J, Puschmann A, Tedgård U, et al. SAMD9 and SAMD9L in inherited predisposition
508 to ataxia, pancytopenia, and myeloid malignancies. *Leukemia*. 2018;32(5):1106–1115.
- 509 45. Inaba T, Honda H, Matsui H. The enigma of monosomy 7. *Blood*. 2018;131(26):2891–2898.
- 510 46. Elghetany MT, Alter BP. p53 protein overexpression in bone marrow biopsies of patients with
511 Shwachman-Diamond syndrome has a prevalence similar to that of patients with refractory
512 anemia. *Arch. Pathol. Lab. Med.* 2002;126(4):452–455.
- 513 47. Sieff CA, Yang J, Merida-Long LB, Lodish HF. Pathogenesis of the erythroid failure in
514 Diamond Blackfan anaemia. *Br. J. Haematol.* 2010;148(4):611–622.
- 515 48. Jaako P, Debnath S, Olsson K, et al. Disruption of the 5S RNP-Mdm2 interaction significantly
516 improves the erythroid defect in a mouse model for Diamond-Blackfan anemia. *Leukemia*.
517 2015;29(11):2221–2229.
- 518 49. Frattini A, Bolamperti S, Valli R, et al. Enhanced p53 Levels Are Involved in the Reduced
519 Mineralization Capacity of Osteoblasts Derived from Shwachman-Diamond Syndrome Subjects.
520 *Int. J. Mol. Sci.* 2021;22(24):13331.
- 521 50. Barlow JL, Drynan LF, Hewett DR, et al. A p53-dependent mechanism underlies macrocytic
522 anemia in a mouse model of human 5q- syndrome. *Nat. Med.* 2010;16(1):59–66.
- 523 51. Turlakis ME, Zhang S, Ball HL, et al. In Vivo Senescence in the Sbds-Deficient Murine
524 Pancreas: Cell-Type Specific Consequences of Translation Insufficiency. *PLoS Genet*.
525 2015;11(6):e1005288.
- 526 52. Sloan KE, Bohnsack MT, Watkins NJ. The 5S RNP couples p53 homeostasis to ribosome
527 biogenesis and nucleolar stress. *Cell Rep.* 2013;5(1):237–247.
- 528 53. Xia J, Miller CA, Baty J, et al. Somatic mutations and clonal hematopoiesis in congenital
529 neutropenia. *Blood*. 2018;131(4):408–416.
- 530 54. Valli R, Minelli A, Galbiati M, et al. Shwachman-Diamond syndrome with clonal interstitial
531 deletion of the long arm of chromosome 20 in bone marrow: haematological features, prognosis
532 and genomic instability. *Br. J. Haematol.* 2019;184(6):974–981.
- 533 55. Lee S, Shin CH, Lee J, et al. Somatic uniparental disomy mitigates the most damaging EFL1
534 allele combination in Shwachman-Diamond syndrome. *Blood*. 2021;138(21):2117–2128.
- 535 56. Machado HE, Øbro NF, Williams N, et al. Convergent somatic evolution commences in utero in
536 a germline ribosomopathy. *Nat. Commun.* 2023;14(1):1–14.
- 537 57. Kennedy SR, Schmitt MW, Fox EJ, et al. Detecting ultralow-frequency mutations by Duplex
538 Sequencing. *Nat. Protoc.* 2014;9(11):2586–2606.
- 539 58. Ip BBK, Wong ATC, Law JHY, et al. Application of droplet digital PCR in minimal residual
540 disease monitoring of rare fusion transcripts and mutations in haematological malignancies. *Sci*.
541 *Rep.* 2024;14(1):6400.
- 542 59. Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for
543 detecting rare and subclonal mutations. *Nat. Rev. Genet.* 2018;19(5):269–285.
- 544 60. Schmitt MW, Kennedy SR, Salk JJ, et al. Detection of ultra-rare mutations by next-generation
545 sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 2012;109(36):14508–14513.
- 546 61. Abascal F, Harvey LMR, Mitchell E, et al. Somatic mutation landscapes at single-molecule
547 resolution. *Nature*. 2021;593(7859):405–410.
- 548 62. Dentre SC, Leshchiner I, Haase K, et al. Characterizing genetic intra-tumor heterogeneity across
549 2,658 human cancer genomes. *Cell*. 2021;184(8):2239–2254.e39.
- 550 63. Dremsek P, Schwarz T, Weil B, et al. Optical Genome Mapping in Routine Human Genetic
551 Diagnostics-Its Advantages and Limitations. *Genes*. 2021;12(12):1958.
- 552 64. Lee-Six H, Øbro NF, Shepherd MS, et al. Population dynamics of normal human blood inferred
553 from somatic mutations. *Nature*. 2018;561(7724):473–478.
- 554 65. Osorio FG, Rosendahl Huber A, Oka R, et al. Somatic Mutations Reveal Lineage Relationships
555 and Age-Related Mutagenesis in Human Hematopoiesis. *Cell Rep.* 2018;25(9):2308–2316.e4.

- 556 66. Welch JS, Ley TJ, Link DC, et al. The origin and evolution of mutations in acute myeloid
557 leukemia. *Cell*. 2012;150(2):264–278.
- 558 67. Mitchell E, Spencer Chapman M, Williams N, et al. Clonal dynamics of haematopoiesis across
559 the human lifespan. *Nature*. 2022;606(7913):343–350.
- 560 68. Williams N, Lee J, Mitchell E, et al. Life histories of myeloproliferative neoplasms inferred from
561 phylogenies. *Nature*. 2022;602(7895):162–168.
- 562 69. Fabre MA, de Almeida JG, Fiorillo E, et al. The longitudinal dynamics and natural history of
563 clonal haematopoiesis. *Nature*. 2022;606(7913):335–342.
- 564 70. Spencer Chapman M, Cull AH, Ciuculescu MF, et al. Clonal selection of hematopoietic stem
565 cells after gene therapy for sickle cell disease. *Nat. Med.* 2023;29(12):3175–3183.
- 566 71. Lee-Six H, Kent DG. Tracking hematopoietic stem cells and their progeny using whole-genome
567 sequencing. *Exp. Hematol.* 2020;83:12–24.
- 568 72. Miller TE, Lareau CA, Verga JA, et al. Mitochondrial variant enrichment from high-throughput
569 single-cell RNA sequencing resolves clonal populations. *Nat. Biotechnol.* 2022;40(7):1030–
570 1034.
- 571 73. Hård J, Mold JE, Eisefeldt J, et al. Long-read whole-genome analysis of human single cells. *Nat.*
572 *Commun.* 2023;14(1):5164.
- 573 74. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat.*
574 *Rev. Genet.* 2016;17(3):175–188.
- 575 75. Lindsley RC, Saber W, Mar BG, et al. Prognostic Mutations in Myelodysplastic Syndrome after
576 Stem-Cell Transplantation. *N. Engl. J. Med.* 2017;376(6):536–547.
- 577 76. Jansko-Gadermeir B, Leisch M, Gassner FJ, et al. Myeloid NGS Analyses of Paired Samples
578 from Bone Marrow and Peripheral Blood Yield Concordant Results: A Prospective Cohort
579 Analysis of the AGMT Study Group. *Cancers*. 2023;15(8):2305.
- 580 77. Beck DB, Ferrada MA, Sikora KA, et al. Somatic Mutations in UBA1 and Severe Adult-Onset
581 Autoinflammatory Disease. *N. Engl. J. Med.* 2020;383(27):2628–2638.
- 582 78. Cesaro S, Pillon M, Sauer M, et al. Long-term outcome after allogeneic hematopoietic stem cell
583 transplantation for Shwachman–Diamond syndrome: a retrospective analysis and a review of the
584 literature by the Severe Aplastic Anemia Working Party of the European Society for Blood and
585 Marrow Transplantation (SAAWP-EBMT). *Bone Marrow Transplant.* 2020;55(9):1796–1809.
- 586 79. Myers K, Hebert K, Antin J, et al. Hematopoietic Stem Cell Transplantation for Shwachman-
587 Diamond Syndrome. *Biol. Blood Marrow Transplant.* 2020;26(8):1446–1451.

588

589 **Figure captions**

590

591 **Figure 1. Using somatic mutations to understand historic HSC clonal relationships.** (A) A
592 simplified phylogenetic tree of HSC relatedness. Based on the mutations detected in each stem cell by
593 WGS, we can build a “family tree” of HSCs using unique and shared somatic mutations. Shapes
594 within the circles represent unique somatic mutations that are inherited by a cell’s progeny. (B) A
595 hypothetical phylogenetic tree. The terminal point of each line represents the stem cell that was
596 sampled. Traveling up these lines, any points of convergence (termed a “coalescence”) indicate an
597 ancestral HSC where two progeny stem cells and their subsequent progeny can be traced. The
598 uppermost branch points represent stem cell divisions that likely occurred in embryogenesis. The
599 dotted box highlights how the structure in (A) would appear within a phylogenetic tree. The presence
600 and distribution of coalescences is related to both aging and disease state. Example tree structures
601 representing HSC relatedness in three different contexts are shown in (C-F). (C) Phylogenies from
602 younger healthy adults appear highly polyclonal with many HSCs contributing to the overall

603 structure,⁶⁷ as indicated by the back of branch points beyond early development. (D) By the time
604 individuals reach their seventh decade of life, tree structures begin to show evidence of decreased
605 clonal diversity,⁶⁷ with a greater number of sampled HSCs being part of large clones. These
606 expansions are often (but not always) associated with non-pathogenic mutations that provide HSCs
607 with a survival advantage. (E) This skewed clonal structure is even more exaggerated in the case of
608 myeloid malignancy. In this context, pathogenic mutations like *JAK2V617F* cause large expansions of
609 specific HSCs. (F) The phylogenetic trees of patients with SDS show evidence of clonal expansions
610 more characteristic of older adults. Colored branches indicate which clones carry driver mutations
611 (trees taken from Machado *et al.*⁵⁶).

612

613 **Figure 2. Information generated using the WGS approach to studying HSC dynamics.** Aside
614 from creating phylogenetic trees using the patterns of shared and unique SNVs generated using the
615 WGS approach, a variety of different information can be mined from this type of experiment. For
616 example, the timing of mutation acquisition can be determined for expanded clones. This potentially
617 has significant implications for when and how myeloid cancer-associated mutations should be
618 monitored and when patients could receive therapeutic intervention. Additionally, overall mutation
619 burdens can be estimated and, along with information on mutational signatures, overall clonality and
620 the range of somatic genetic rescue mutations, these data can help to build a more comprehensive
621 picture of the pathways and genes involved in disease-related clonal dynamics. Importantly, the
622 source of cellular material (bone marrow vs. peripheral blood) was not found to significantly
623 influence these results. Lee-Six *et al.*⁶⁴ sorted single Lin⁻CD34⁺CD38⁻CD90⁻CD45RA⁺ HSCs and bulk
624 HSPCs (MEPs, GMPs, CMPs) from both peripheral blood and bone marrow of the same individual
625 and did not observe any significant differences in mutational burden between the two compartments.
626 Also, the phylogenetic tree built exclusively from one source versus the other was not significantly
627 different. This is because sampled HSCs and HSPCs all report on their parent HSCs.

628

629 **Acknowledgements**

630

631 Work in the D.G.K. laboratory is supported by the Bill and Melinda Gates Foundation (INV-002189
632 and INV-038816, a Cancer Research UK Programme Foundation Award (DCRPGF\100008), the UK
633 Medical Research Council (MC_PC_21043, MR/V005502/1, MR/Y011945/1), and the National
634 Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC)
635 (NIHR203331). Work in the AJW laboratory is supported by the UK MRC (MR/T012412/1), Blood
636 Cancer UK (21002), Rosetrees Trust (PGL22/100032), Isaac Newton Trust (G125517),
637 Addenbrookes Charitable Trust (900426), SDSUK, SDS Foundation, Butterfly Guild, Connor Wright
638 Project, the Cambridge National Institute for Health Research Biomedical Research Centre, European

639 COST Actions CA18233 and CA21154.

640

641 **Authorship and conflict-of-interest statements**

642

643 A.H.C., D.G.K. and A.J.W. contributed equally to the text of this paper and A.H.C. created the
644 figures. A.J.W. is a consultant for SDS Therapeutics. The remaining authors declare no competing
645 interests.